PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/00

A1

(11) International Publication Number: WO 00/08568

(43) International Publication Date: 17 February 2000 (17.02.00)

(21) International Application Number: PCT/US99/17655

(22) International Filing Date: 4 August 1999 (04.08.99)

(30) Priority Data:

60/095,308	4 August 1998 (04.08.98)	US
09/282,392	31 March 1999 (31.03.99)	US

(71) Applicant: DRYKEN TECHNOLOGIES [US/US]; 2800 Industrial Terrace Drive, Austin, TX 78759 (US).

(72) Inventors: VANDERVELDT, Ingrid; 8541 Capital of Texas Highway # 3082, Austin, TX 78759 (US). BLACK, Christopher, L.; 131 Northshore Drive, Andersonville, TN 37705 (US).

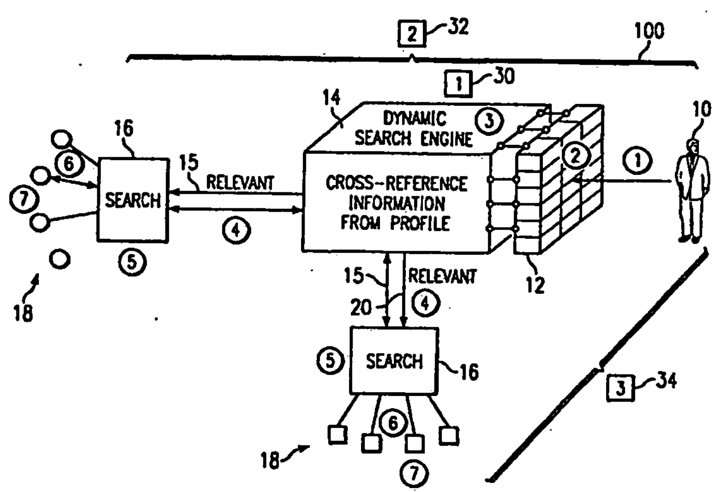
(74) Agent: McLAUCHLAN, Robert, A.; Gray Cary Ware & Freidenrich LLP, Suite 1440, 100 Congress Avenue, Austin, TX 78701 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

(54) Title: METHOD AND SYSTEM FOR DYNAMIC DATA-MINING AND ON-LINE COMMUNICATION OF CUSTOMIZED INFORMATION



(57) Abstract

The present invention provides a method and system for dynamically searching databases in response to a query, and more specifically, a system and method for dynamic data-mining and on-line communication of customized information. This method includes the step of first creating a search-specific profile (15). This search-specific profile is then input into a data-mining search engine (100). The data-mining search engine will mine the search-specific profile to determine topics of interests. These topics of interest are output to at least one search tool (16). These search tools (16) match the topics of interest to at least one destination data site wherein the destination data sites are evaluated to determine if relevant information is present in the destination data site. Relevant information is filtered and presented to the user (10) making the inquiry.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

METHOD AND SYSTEM FOR DYNAMIC DATA-MINING AND ON-LINE COMMUNICATION OF CUSTOMIZED INFORMATION

RELATED APPLICATIONS

This application claims benefit of U.S. Provisional Application No. 60/095,308 filed on August 4, 1998. Additionally this application incorporates by reference the prior U.S. Provisional Application No. 60/095,308 filed on August 4, 1998 entitled "Method and System for Dynamic Data-mining and On-line Communication of Customized Information" to Ingrid Vanderveldt and U.S. Patent Application No. 09/282,392 filed on March 31, 1999 entitled "An Improved Method and System for Training an Artificial Neural Network" to Christopher L. Black.

TECHNICAL FIELD OF THE INVENTION

This invention relates generally to the use of a dynamic search engine and, more particularly, to a dynamic search engine applied to the Internet that allows for customized queries and relevant responses.

BACKGROUND OF THE INVENTION

Current Internet search tools often provide irrelevant data sites or web sites. Often, current search tools provide a score of relevance according to text frequency within a given data site or web page. For example, "termites" and "Tasmania" and "not apples":

- If a web page has several instances of the word "termites" (600 for example), the web page would receive a high relevance score.

- A web page with 600 "termites" and one "Tasmania" would receive a slightly higher score.
- A web page with the above plus "apples" would then receive a slightly lesser score.

Therefore, a score of relevance according to a data site or web page is often based on text or word frequency. Therefore current search tools often provide a list of irrelevant web pages. Furthermore, there is the opportunity for abuse in and associated with the method of the available search tools. Current search tools often provide links that are stale (old data that is no longer at the address of the data site). Existing search tools utilize indices that are compiled in the background continuously. However, with respect to an individual query, a historical result is received. Therefore, the search process involves a large amount of filtering by the individual user.

Therefore, there is a need to more efficiently utilize search tools to overcome irrelevant results. At present, it is desirable to have an efficient method for performing a search which would take into account demographic as well as historical user information to filter irrelevant data from the results from existing search tools.

Furthermore, it is desirable to have a search engine which will evaluate and filter stale data responses from an existing search tool response.

SUMMARY OF THE INVENTION

In accordance with the present invention, a method and system for searching databases in response to a query is provided that substantially eliminates or reduces disadvantages and problems associated with previous methods and systems for searching databases.

More specifically, the present invention provides a system and method for dynamic data-mining and on-line communication of customized information. This method includes the steps of first creating a search-specific profile. This search-specific profile is then inputted into a data-mining search engine. The data-mining search engine will mine the search-specific profile to determine at least one topic of interest. The at least one topic of interest may comprise a specific and/or related topics to interest. The at least one topic of interest is outputted to at least one search tool. These search tools match the at least one topic of interest to at least one destination data site. The destination data sites (web page) are evaluated to determine if relevant information is present in the destination data site. If relevant information is present at the destination data site, this data site may be presented to a user.

One broad aspect of the present invention includes the coupling of a data-mining search engine to at least one search tool. This data-mining search engine can review and evaluate data sites. Current search tools available may create a massive index of potential data sites. The data-mining engine of the present invention evaluates whether

data accumulated by current search tools are relevant to a user and filters out non-relevant information.

The present invention provides an advantage by providing a search engine algorithm that provides fresh (as opposed to stale) links to more highly relevant web pages (data sites) than provided by the current search engines.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference is now made to the following description taken in conjunction with the accompanying drawings in which like reference numerals indicate like features and wherein:

FIGURE 1 shows a diagram of the present embodiment of the invention;

FIGURE 2 illustrates an example of operating the present invention;

FIGURE 3 explains the related patent applications to the present invention;

FIGURE 4 depicts the use of a training scheme according to the teachings of BLACK; and

FIGURE 5 details a flow chart illustrating the method of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Preferred embodiments of the present invention are illustrated in the FIGUREs, like numerals being used to refer to like and corresponding parts of the various drawings.

In accordance with the present invention, a method and system for dynamically searching databases in response to a query is provided that substantially eliminates or reduces disadvantages and problems associated with previous methods and systems for searching databases.

More specifically, the present invention provides a system and method for dynamic data-mining and on-line communication of customized information. This method includes the steps of first creating a search-specific profile. This search-specific profile is then inputted into a data-mining search engine. The data-mining search engine will mine the search-specific profile to determine at least one topic of interest. The at least one topic of interest may comprise a specific and/or related topics to interest. The topic of interest is outputted to at least one search tool. These search tools match the topic of interest to at least one destination data site. The destination data sites are evaluated to determine if relevant information is present in the destination data site. If relevant information is present, this data site is assigned a relevance score and presented to user requesting the query.

One broad aspect of the present invention includes the coupling of a data-mining search engine to at least one search tool. This data-mining search engine reviews and

evaluates available data and data sites. Current search tools available may create a massive index of potential data sites. The data-mining engine of the present invention evaluates whether the available data accumulated by current search tools are relevant to a user and filters out all non-relevant information, creating a more effective and efficient search engine.

In one embodiment, the present invention includes a web site containing several data-mining tools. These tools fall into two separate categories: a dynamic approach to generating a list of links that are well correlated to a user provided search string using a novel search strategy (e.g., incorporating simple text matching, text associations, synonym and near text matching - to handle misspellings, profile information, a recursive definition of document importance/relevance – important/relevant documents link to other important/relevant – and weighting of the previous factors based upon Al), and stand-alone models (e.g., neural networks and NSET models, as well as others known to those skilled in the art), which would provide useful predictions or estimations (such as described in the U.S. Patent Application No. 09/282,392 entitled "An Improved Method and System for Training An Artificial Neural Network" filed 31 March 1999 to Christopher L. Black, hereafter BLACK.

The stand alone models would be created with implementer or user interaction, and could be ever increase in number, as desired and as data was discovered/licensed/acquired. Eventually, the web site would contain a portal to hundreds of thousands of

interesting and useful models.

Neither the search engine nor the models would necessarily be limited to medical information and topics. However, the present invention primarily focuses on healthcare-related applications. The system and method of the present invention need not be limited to such health care database.

The present invention provides a method for data-mining that provides use of many different Al models derived for many different applications from many different datasets. The present invention provides the benefit of a neural network training algorithm, genetic algorithms expert and fuzzy logic systems, decision trees, and other methods known to those skilled in the art applied to any available data.

Secondly, the present invention allows the compact storage, retrieval, and use of relationships and patterns present in many datasets, each made up of very many patterns of examples, each made of several different measurements or values, each requiring several bytes when stored conventionally or explicitly (as in a relational database or a flat file). Single datasets consisting of multiple gigabytes and terabytes of data are routinely being generated, with exabyte datasets looming on the horizon. With the use of multiple modeling techniques (different approaches are appropriate to different applications), models encapsulating and summarizing useful information contained within hundreds or even thousands of these datasets could stored on a single consumer level personal computer hard drive.

FIGURE 1 illustrates one physical implementation of the present invention. The number of servers, interconnections, software modules, and the like would largely be determined by scalability concerns. The web site 12 would consist of a graphical user interface (GUI) to present dynamically generated indexes and forms that allow the user 10 to provide a search profile and submit their search requests or feed inputs into a selected Al model. The web site 12 could reside upon a single or on a standard farm of web server machines. Search engine requests 15 would be provided to a single or a farm of search machines 16, which would either query a static public or proprietary databases 18/indices of links either pre-created (and continually updated) or licensed from, for example, Yahoo and other link search engines. This static list (formed from data sites 18) would provide a starting point for a dynamic (live) search. Both search machines/machine farms 16 would require extremely high speed access to the Internet or other like data networks.

Data-mining is the process of discovering useful patterns and relationships within data. This is typically accomplished by training and then applying a neural network, or inducing and then applying a decision tree, or applying a genetic algorithm, etc. Once the training aspect of many of the techniques is performed, the result is the data-mining tool (e.g., a trained neural network – into which someone who knows nothing about Al can simply input values and receive results).

Data-mining "tools" are discrete and specific. Certain models are appropriate for certain tasks. When

explanation of a particular result is important (as in credit approval/rejections), and the available data supports the generation/formulation of rules, an expert or fuzzy logic system might be appropriate. When optimization of a particular quantity is important, a genetic algorithm or another evolutionary algorithm might be more useful. When prediction/estimation is important, the neural network training algorithm might be used.

The Dynamic Search Engine 100 can extract/provide useful information from publicly and freely available databases 18. However, the present invention can do the same with proprietary databases 18.

One embodiment of the present invention incorporates an enhanced version of simple text matching (allowing reduced weight for synonym and possible misspelling matches) at the first level. Associations with profile information provides a second metric of relevance (e.g., certain words and word combinations are found to correlate with interest for people providing certain combinations of search profile factors). The final metric is whether other articles possessing high (normalized) relevance (using all 3 levels – a recursive definition) link to the page in question. If so, then the relevance as established by this metric is high.

The spidering/crawling/roboting starts from the static index found in response to the initial query 15 of databases 18. Data sites included in the index are scanned and assigned relevance using the 3 facts above. Data Sites with high levels of relevance are scanned deeper (a links are followed, as well as the links revealed on those

subsequent pages) than non-relevant pages. After a maximum number of links have been followed, or the total relevance of pages indexed exceeds a threshold, the search stops and results 20 are returned to user 10, organized by a weighted conglomeration of the 3 factors (generated by a neural network trained upon the user profile and previous searches and relevance results).

For the pre-created models, the present invention also has a page indexing the available canned models that the user could simply choose from. Alternatively, based upon text entered at the dynamic search engine GUI 12, the dynamic search engine could suggest appropriate models, where appropriate (e.g., if user enters blue book, the present invention could return at the top of a list of links, a link to a used car value estimator neural network).

FIGURE 2 illustrates one embodiment of the present invention wherein the search tools comprise a privately licensed search tool 22 accessing privately held databases 24 and publicly available database 18 accessed by search tools provided by YAHOO, EXCISE, LYCOS and other search tools known to those skilled in the art.

FIGURE 3 provides an overall description of three processes which occur within Figures 1 and 2. Process 30 illustrates the dynamic search engine application which performs the function of mining search profile data as provided from user 10 via GUI 12. Mining or cross referencing the search profile data against subject information includes the dynamic search capabilities of evaluating data sites 18. Process 32 in Figure 1

illustrates the interaction between a user 10, the dynamic search engine and an available search tool 16, which accesses individual web sites 18. Search tool 16 for each individual may be customized to the protocols associated with each search engine. Process 34 illustrates the process between a user 10, a dynamic search engine of the present invention and a proprietary search engine when the search tool 16 is a proprietary search engine accessing proprietary databases.

The improvements to previously existing artificial neural network training methods and systems mentioned in the various embodiments of this invention can occur in the conjunction with one another (sometimes even to address the same problem). FIGURE 4 demonstrates one way in which the various embodiments of an improved method for training an artificial neural network (ANN) can be implemented and scheduled. FIGURE 4 does not demonstrate how representative dataset selection is accomplished, but instead starts at train net block 101 with representative training dataset already selected.

The training dataset at block 101 can consist initially of one kind of pattern that is randomly selected, depending on whether or not clustering is used. Where clustering takes place, it takes place prior to any other data selection. Assuming, as an example, that clustering has been employed to select twenty training patterns, ANN can then be randomly initialized, all the parameters can be randomly initialized around zero, and ANN can take those 20 data patterns and for each one calculate the gradient and multiply the gradient by the initial value of the learning

rate. The adaptive learning rate is user-definable, but is usually initially set around unity (1). For each of the representative data patterns initially selected, the training algorithm of this invention calculates the incremental weight step, and after it has been presented all twenty of the data patterns, it will take the sum of all those weight steps. All of the above occurs at train net block 101.

From train net block 101, the training algorithm of this invention goes to step 102 and determines whether the training algorithm is stuck. Being stuck means that the training algorithm took too large a step and the prediction error increased. Once the training algorithm determines that it is stuck at block 104 it decreases the adaptive learning rate by multiplying it by a user-specified value. A typical value is 0.8, which decreases the learning rate by 20%.

If the training algorithm reaches block 102 and determines there has been a decrease in the prediction error (i.e., it is not stuck), the training algorithm proceeds to block 108 and increases the learning rate. The training algorithm returns to block 101 from block 108 to continue training the ANN with a now increased adaptive learning rate.

The training algorithm proceeds to block 106 after decreasing the adaptive learning rate in block 104 and determines whether it has become "really stuck." "Really stuck" means that the adaptive learning rate decreased to some absurdly small value on the order of 10^{-6}. Such a reduction in the adaptive learning rate can come about as a

result of the training algorithm landing in a local minimum in the error surface. The adaptive learning rate will normally attempt to wiggle through whatever fine details are on the error surface to come to a smaller error point However, in the natural concavity or flat spot of a local minimum there is no such finer detail that the training algorithm can wiggle down to. In such a case the adaptive learning rate decreases to an absurdly low number.

If at block 106, if the training algorithm determines that it is really stuck (i.e., that the learning rate has iteratively decreased to an absurdly small value), it proceeds to block 110 and resets the adaptive learning rate to its default initial value. In the event that the training algorithm is <u>not</u> really stuck at block 106, it returns to block 101, recalculates the weight steps, and continues training with newly-modified weights. The training algorithm continues through the flow diagram, as discussed above and below.

Once the adaptive learning rate is reset at block 110, the training algorithm proceeds to block 112, where it determines whether the minimum in which it is currently stuck is the same minimum in which it has been stuck in the past (if it has been stuck before). This is because as the training algorithm is learning it will sometimes get out of a local minimum and wind up in the same minima at a future time. If it finds itself stuck in the same minimum, the training algorithm checks, at block 114, whether it has achieved a maximum on the gaussian distribution from which a random value is chosen to perturb the weights (i.e., whether the maximum jog strength has been achieved). The

"maximum jog strength" is the maximum value from the gaussian distribution. If the maximum jog strength has been achieved, at block 116 the training algorithm resets the jogging strength.

The jogging strength is reset at block 116 because the problem is not so much that the training algorithm has found itself in a local minimum, but that the ANN is not complicated enough. The training algorithm moves to block 118 and determines whether it has, prior to this point, trimmed any weights. "Trimming weights" means to set those weights to zero and take them out of the training algorithm. The procedure for trimming of weights will be described more fully with respect to FIGURE 13 below.

If at step 118 the training algorithm determines that weights have previously been trimmed (i.e., that the weights have been previously randomly affected but the training algorithm still wound up in the same minimum because the network was not complex enough to get any more accuracy out of the mapping), the training algorithm moves to step 120 and untrims 5% of the weights. This means that weights that were previously trimmed are allowed to resume at their previous value, and from this point on they will take part in the training algorithm. The training algorithm returns to step 101 and continues to train as before.

By untrimming 5% of the weights, the training algorithm returns a little more complexity back to the model in hopes of decreasing the prediction error. If prediction error does not decrease, the training algorithm will once again reach a local minimum and the training

algorithm will determine once again at block 112 whether it is stuck in the same minimum as before. Note, however, that at block 110 the adaptive learning rate is reset before addressing the complexity issue of untrimming previously trimmed weights, so it takes some iterations through blocks 101, 102, 104, 106 and 110 before getting back to the process of untrimming any more weights. In the event the training algorithm does wind up in the same minimum, the maximum jog strength will not have been reached, since it was previously reset at block 116 in a prior iteration. Instead, the training algorithm will proceed to block 136. At block 136 the weights are jogged, and at block 140 the jogging strength is slightly increased according to a gaussian distribution. Following block 140, the training algorithm proceeds to train net block 101 and continues training.

If in the course of training the training algorithm again reaches the same minimum, the procedure above is repeated. In the event the jog strength once again reaches the maximum level at block 114, the training algorithm resets the jogging strength as previously discussed. If the training algorithm reaches block 118 after several rounds of untrimming weights that there are no longer any trimmed weights, the training algorithm proceeds along the "no" path to block 122.

At block 122, the training algorithm determines if this is the first time it has maxed out the jog strength on this size ANN. The training algorithm keeps a counter of how many times the jog strength has maxed out with an ANN of a given size. If this is the fist time the jog strength

has maxed out for the current ANN size, the training algorithm proceeds along the "yes" path to block 124 and completely re-initializes the ANN. All of the weights are re-initialized and the ANN is restarted from scratch. The training algorithm proceeds to block 101 and commences training the net anew. The ANN, however, remains whatever size it was in terms of number of hidden layers and number of nodes when training resumes at train net block 101 with the newly re-initialized weights.

At block 122, if the answer is "no," the training algorithm proceeds along the "no" path to block 126. At block 126 the training algorithm has already maxed out the jog strength more than once for the current size ANN. Block 126 tests to see how many new nodes have been added for the current state of the representative training dataset. The training algorithm determines if the number of new nodes added for this size ANN is greater than or equal to five times the number of hidden layers in the ANN. If the number of new nodes added is not equal to or in excess of 5 times the number of hidden layers in the ANN, the training algorithm proceeds along the "no" path to block 128. At block 128, a new node is added according to the procedures discussed above and the training algorithm proceeds to train net block 101 to continue training the artificial neural network with the addition of the new node. The training algorithm of this invention will then proceed as discussed above.

If the number of new nodes added exceeds five times the number of hidden layers, the training algorithm proceeds along the "yes" path from block 126 to block 130.

At block 130, the training algorithm determines whether a new layer has previously been added to the ANN. If the training algorithm has not previously added a new layer (since the last time it added a training data pattern), it proceeds along the "no" path to block 132 and adds a new layer to the artificial neural network. The training algorithm then proceeds to block 101 and continues to train the net with the newly added layer. If a new layer has been added since the last training pattern was added, the training algorithm proceeds along the "yes" path to block 134.

If a new layer has previously been added, it means that the training algorithm has previously added a number of nodes, has jogged the weights a number of times, and has added a layer because of the new training data pattern that has been added in the previous iteration. The training algorithm decides by going to block 134 that the training data pattern added recently is an out-lier and does not fit in with the other patterns that the neural network recognizes. In such a case, at block 134 the training algorithm removes that training data pattern from the representative training dataset and also removes it from the larger pool of data records from which the training algorithm is automatically selecting the training dataset. The training algorithm once again proceeds to train net block 101 and continues to train the network without the deleted data pattern.

Returning to block 112, if the training algorithm decides that it has not fallen into the same minimum, it proceeds along the "no" path to block 138. At block 138,

the training algorithm resets the jogging strength to give only a small random pèrturbation to the weights and parameters in an attempt to extricate itself from a new local minimum. If the training algorithm reaches a new local minima, we want the training algorithm to start over again. It is desirable to reset the jogging strength because to give a small random perturbation to the weights and parameters. The intent is to start off with a small perturbation and see if it is sufficient to extricate the training algorithm from the new local minimum.

After resetting the jogging strength in block 138, the training algorithm proceeds to block 136 and jogs the weights. The training algorithm proceeds to block 140, increases the jogging strength, and proceeds to block 101 and trains the net with the newly increased jogging strength.

FIGURE 4 thus gives us an overview in operation of the various embodiments of the training algorithm of BLACK.

FIGURE 5 provides a flow chart of the present invention illustrating one method of dynamic data-mining.

At step 202, user 10 arrives at a GUI 12 and logs on. Once logged in, the system queries the user for their specific search profile.

Once the user has entered the data, the specific profile is output to data-mining search engine 12 at step 204.

In step 206, the dynamic search engine 100, data mines the specific profile to determine what other related topics of interest would be relevant and of greatest to the user 10.

The information is categorized so that it can be transferred to both existing and future search engines.

These related topics of interest are fed back to user 10. In step 208 user 10 then determines the topic outputs the specific and related topics to be researched. The dynamic search engine then connects existing public and proprietary search tools 16.

At step 210, the information is transferred, over the Internet, or other like communication pathway, to other sites and/or licensed search tools (Yahoo, Lycos or others known to those skilled in the art) to find matching the search query 15.

At step 212, information is gathered from the search destination site(s) pertaining to the request.

At step 214, information is sent, from the search engine (Yahoo, etc.) to the dynamic search engine. Relevant information is gathered from the destination databases.

The information is sent back to the data-mining search engine 14 at which point the information is cross-referenced to the user's profile. Depending on the profile, the presentation will rate, weigh and organize each search to present the most relevant and related topics of interest.

The information will be presented back to the user in a way such as:

- The most relevant topics/areas of interest: #1-10
- The most related topics/area of Interest: #1-10

This information will include subjects such as areas of interest that have shown to have a strong correlation and/or relationship to the specific topic of interest.

Once the user has received the information, they will be asked if they would like to see more information. Each time the user requests additional information, it will be presented in subsequent to the most recent, most relevant, information previously presented.

Over time, the profile information database will continue to grow and become more intelligent. Therefore, each subsequent searches will become more intelligent and relevant to the previous user. This data will continue to collect in a profile database located within Dynamic search engine 14. Over time, one can monitor the searches, and rate each search a success or failure (or some degree of one or the other), to then optimize with Artificial Neural Nets and Genetic algorithms, or other empirical techniques used in conducting the search.

The Dynamic search engine becomes an intelligent agent that specifically pulls back better (and more recent – also implying more thorough) results than the static search engines that require more user information. Results are specifically searched for with user needs expressed prior to the search. Resulting in explicitly tailored searches to a user request.

One embodiment of the present invention provides for a multi-component tool, with six main interacting components – Web servers, Highspeed Internet Connections, Web pages, Health-related Databases, Database Query and Responses Scripts/Code, and the Dynamic Internet Search Scripts/Code.

The web servers are the computer equipment, operating systems, and communications software that will contain and execute the web pages, (GUI) 2 and Dynamic search engine 14. The equipment may also contain the databases, provide highspeed Internet connections, and perform the database 18 and Internet searches. This equipment may be configured from off-the-shelf workstations, peripherals, and software. Initially, only one system must be configured. However, as use grows a search response-time per user can be estimated (and a scalability strategy developed). This will enable projection of the number of servers necessary per user. Estimates may be arrived from data provided by similar web service companies.

The communication pathways, Highspeed Internet connections, consist of T1s, T3s, or other connections known to those skilled in the art. Those connections provide wide-bandwidth communication to and from the entire Internet, and any associated equipment which is not considered a part of the web server. As with the web servers, the amount of necessary bandwidth will be a function of number of concurrent users.

Web pages (GUI) 12 present search prompts and results via the Internet to user 10 and define the interface to the system of the present invention to the user.

The web pages define the format of the query pages and search result pages. The query pages must have multiple forms/options to allow flexibility in searching (which databases to query, simple/Boolean forms, whether to search the Internet, how deep/long to search the Internet, etc.). The search result pages will take multiple forms, depending

on the specified request, but will include relevance scores, titles and links, and summaries, much as resulting from internet search engine requests. For internet search results, links would lead to web pages. For other database results, the links would lead to graphical/textual reports for each "hit."

The present invention may utilize databases containing licensed and public domain. This component includes only bare-data and "pre-processing" thereof. Data-mining (e.g., a hypothetical diagnostic tool "what illness you probably have" based upon a neural network trained from a symptom/illness database) and analysis are considered part of the following component and its development.

The database query scripts direct the simple searching and querying of the databases, accesses custom data-mining solutions developed for some of the databases, and allows visualization for exploration of the databases. These scripts are also responsible for returning the results of searches in the HTML format design.

Each data-mining tool to be implemented may be custom developed for the appropriate database. Such tools will continue to be added, as appropriate data becomes available to the present invention, even after deployment of the system.

These scripts, based upon the text-based query, and possibly a demographic and historical search profile, perform a "blind" an "dynamic" search of world wide web pages, returning those deemed most "relevant." This search is blind, in that prior to the search, no index (such as those compiled and used by existing search engines) has

been generated. This search will be dynamic, in that contrary to the manner in which other search engines return their results (based upon a pre-compiled though continuously updated index) the web is searched anew with each request.

Based upon the top N (adjustable by the user) results returned by the static search, the dynamic search would assign a relevance to each page. The dynamic search would then proceed to "spider" to each of the links contained in each page, according to a function of the relevance. The search would spider several levels beyond extremely relevant pages, and none beyond irrelevant pages. As listed below, initially the relevance function would consist of simple text matching and counting of keyword occurrences (as do the other search engines).

Based upon a historical profile of search successes and failures as well as demographic/personal data, technologies from artificial intelligence and other fields will optimize the relevance rating function. The more the tool is used (especially by a particular user) the better it will function at obtaining the desired information earlier in a search. The user will not have to be a computer or information scientist. The user will just be aware that with the same input the user might give a static search engine, the present invention finds more relevant, more recent and more thorough results than any other search engines.

A method and system for dynamically searching databases in response to a query is provided by the present invention. More specifically, a system and method for

dynamic data-mining and on-line communication of customized information. This method includes the steps of first creating a search-specific profile. This search-specific profile is then inputted into a data-mining search engine. The data-mining search engine will mine the search-specific profile to determine topic of interests. These topics of interest are outputted to at least one search tool. These search tools match the topics of interest to at least one destination data site wherein the destination data sites are evaluated to determine if relevant information is present in the destination data site. Relevant information is filtered and presented to the user making the inquiry.

The present invention provides an advantage by providing a search engine algorithm that provides fresh (as opposed to stale) links to more highly relevant web pages (data sites) than provided by the current search engines.

Although the present invention has been described in detail, it should be understood that various changes, substitutions and alterations can be made hereto without departing from the spirit and scope of the invention as described by the appended claims.

WHAT IS CLAIMED IS:

1. A method of dynamically searching databases in response to a query, comprising the steps of:

profiling a user to create a user-specific profile;

inputting said user-specific profile to a data-mining search engine;

mining said user-specific profile to determine at least one topic of interest;

outputting said at least one topic of interest to at least one search tool;

using said at least one search tool to match said at least one topic of interest to at least one destination data site;

evaluating said at least one destination data site for relevant information; and

presenting said relevant information to said user.

2. The method of Claim 1, wherein said at least one topic of interest further comprises specific and related topics of interest.

3. A dynamic search engine comprising:

a server system;

a software program executed on said server system wherein said software program is operable to provide a graphical user interface to a user in which a search query may be received;

a data-mining engine operable to receive said search query;

at least one search tool coupled to said data-mining engine operable to execute said search query and receive a response; and

a filtering system to evaluate said response and pass relevant response data from said response to said user.

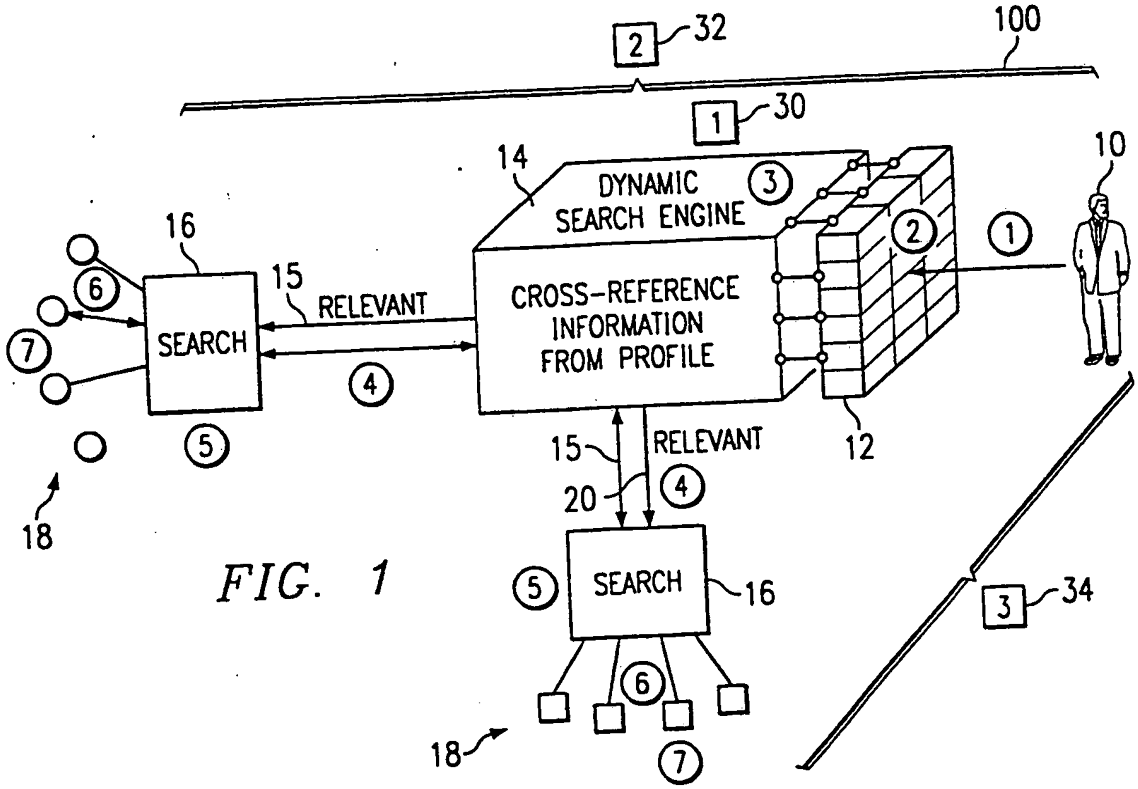

2
32
100
1
30
14
DYNAMIC SEARCH ENGINE
3
2
1
10
16
6
15
RELEVANT
SEARCH
7
CROSS-REFERENCE INFORMATION FROM PROFILE
4
5
15
RELEVANT
12
20
4
18
SEARCH
5
16
3
34
6
18
7

FIG. 1

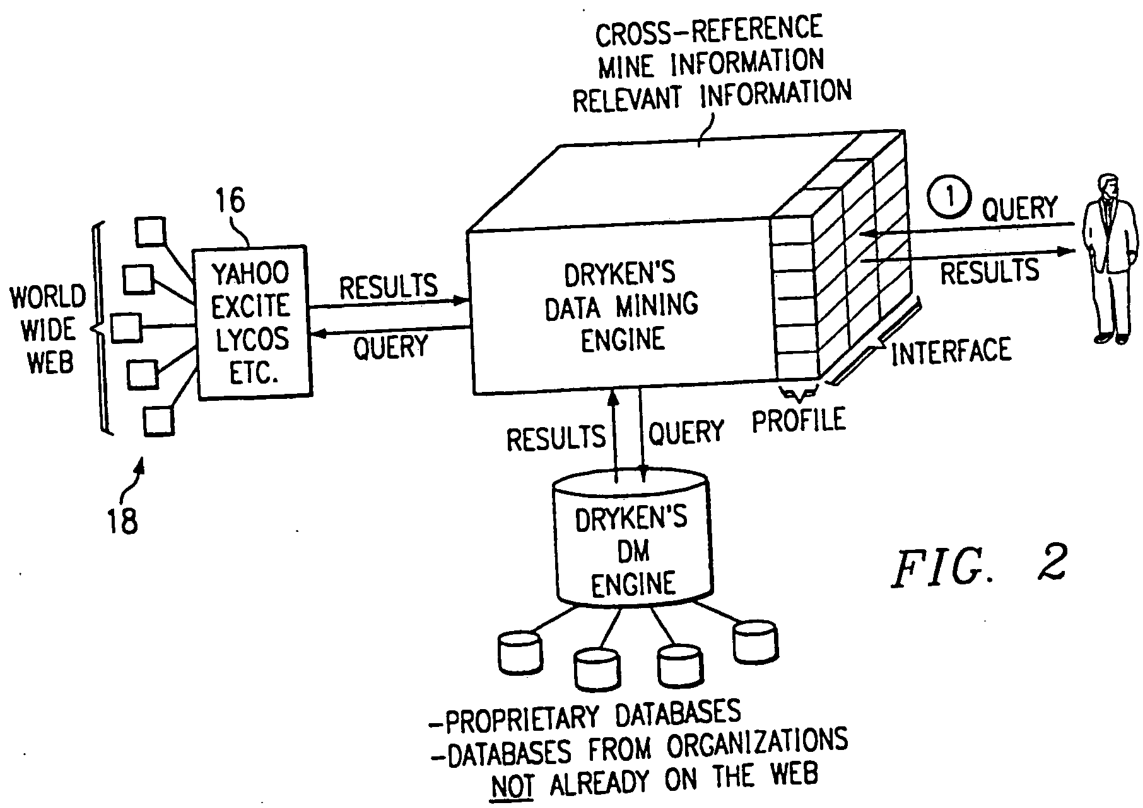

CROSS-REFERENCE MINE INFORMATION RELEVANT INFORMATION
16
1 QUERY
WORLD WIDE WEB
YAHOO EXCITE LYCOS ETC.
RESULTS
QUERY
DRYKEN'S DATA MINING ENGINE
RESULTS
INTERFACE
RESULTS
QUERY
PROFILE
18
DRYKEN'S DM ENGINE
-PROPRIETARY DATABASES
-DATABASES FROM ORGANIZATIONS NOT ALREADY ON THE WEB

FIG. 2

PROCESS INFORMATION

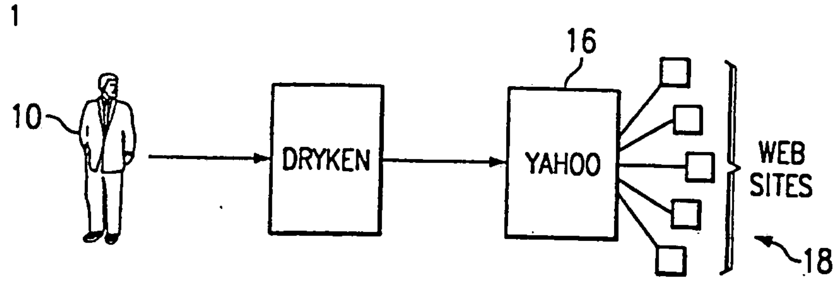
1) THE ENTIRE PROCESS

1 IN FIG. 1

2) BOX #3 APPLICATION
- -MINING PROFILE DATA
- -MINING (CROSS-REFERENCING) PROFILE DATA AGAINST SUBJECT INFORMATION
- -DYNAMIC SEARCH CAPABILITIES

2 IN FIG. 1

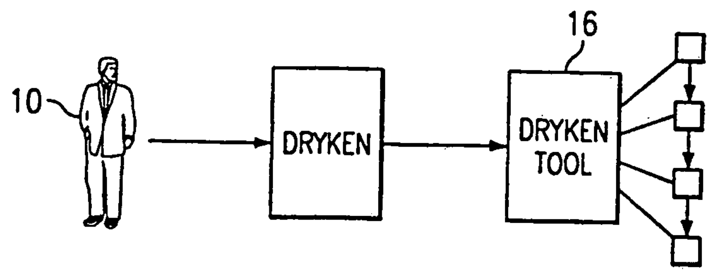
3) THE PROCESS BETWEEN USER AND WWW SEARCH TOOL



4) BOX #5 APPLICATION
- -DATA MINING SPECIFIC DATABASE INFORMATION

3 IN FIG. 1

5) THE PROCESS BETWEEN USER AND PROPRIETARY OR OTHER NON-READILY ACCESSIBLE DATABASES



CB. • 6) THE DATA MINING TECHNOLOGY ITSELF

• 7) THE DYNAMIC TRAINING ALGORITHM
-NEW METHOD

FIG. 3

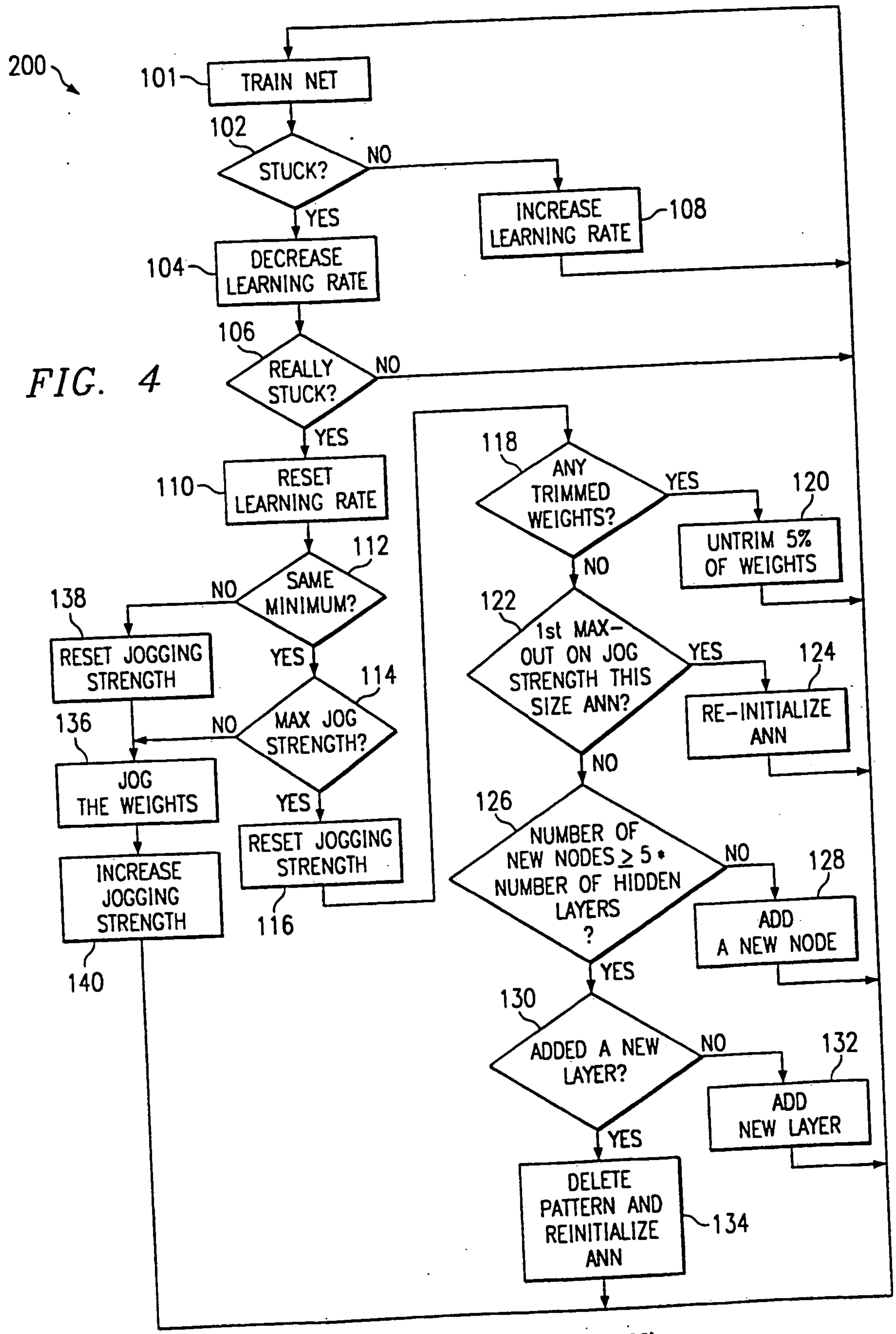

200
101
TRAIN NET
102
STUCK?
NO
YES
INCREASE LEARNING RATE
108
104
DECREASE LEARNING RATE
106
REALLY STUCK?
NO
YES
110
RESET LEARNING RATE
112
SAME MINIMUM?
NO
YES
138
RESET JOGGING STRENGTH
114
MAX JOG STRENGTH?
NO
YES
136
JOG THE WEIGHTS
RESET JOGGING STRENGTH
116
INCREASE JOGGING STRENGTH
140
118
ANY TRIMMED WEIGHTS?
YES
NO
120
UNTRIM 5% OF WEIGHTS
122
1st MAX-OUT ON JOG STRENGTH THIS SIZE ANN?
YES
NO
124
RE-INITIALIZE ANN
126
NUMBER OF NEW NODES ≥ 5 * NUMBER OF HIDDEN LAYERS ?
NO
YES
128
ADD A NEW NODE
130
ADDED A NEW LAYER?
NO
YES
132
ADD NEW LAYER
DELETE PATTERN AND REINITIALIZE ANN
134

FIG. 4

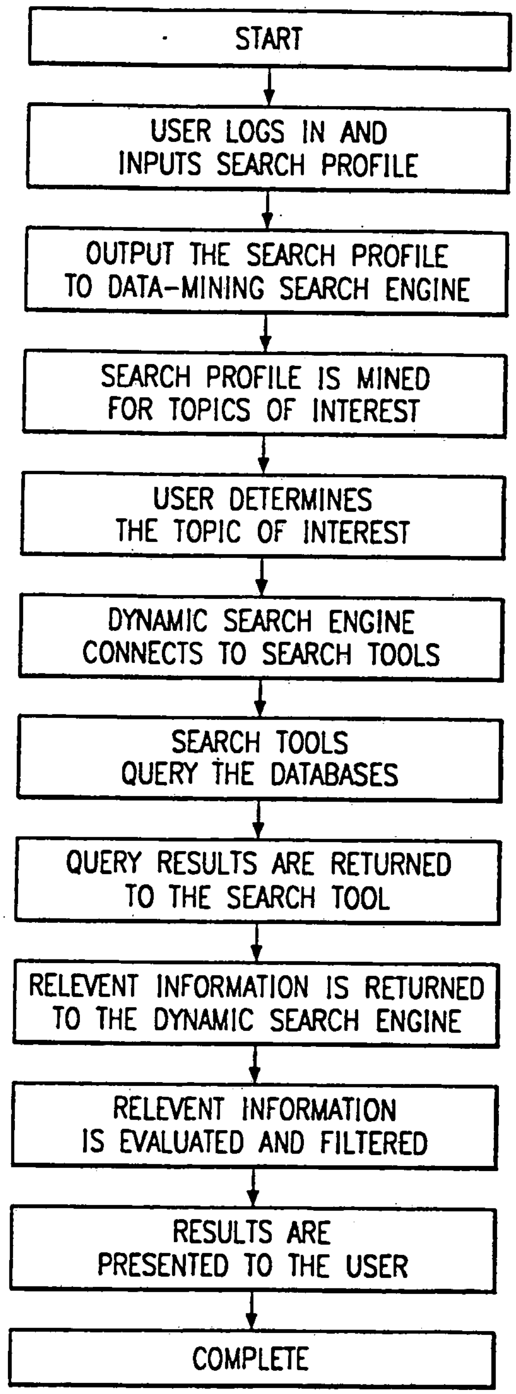

START
USER LOGS IN AND INPUTS SEARCH PROFILE
OUTPUT THE SEARCH PROFILE TO DATA-MINING SEARCH ENGINE
SEARCH PROFILE IS MINED FOR TOPICS OF INTEREST
USER DETERMINES THE TOPIC OF INTEREST
DYNAMIC SEARCH ENGINE CONNECTS TO SEARCH TOOLS
SEARCH TOOLS QUERY THE DATABASES
QUERY RESULTS ARE RETURNED TO THE SEARCH TOOL
RELEVENT INFORMATION IS RETURNED TO THE DYNAMIC SEARCH ENGINE
RELEVENT INFORMATION IS EVALUATED AND FILTERED
RESULTS ARE PRESENTED TO THE USER
COMPLETE

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/17655

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : GO6F 17/00
US CL : 707/5, 10
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/5, 10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Alta Vista, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,933,827 A (COLE et al.) 03 August 1999, col 1, line 56 to col. 2, line 56.	1-3
X	US 5,787,424 A (HILL et al.) 28 July 1998, col. 2, line 47 to col. 3, line 59.	1-3
X	US 5,765,028 A (GLADDEN et al.) 09 June 1998, col. 5, lines 3-65.	1-3
X	US 5,761,662 A (DASAN) 02 June 1998, col. 6, line 20 to col. 8, line 51.	1-3
X	US 5,634,051 A (THOMSON) 27 May 1997, col. 6, line 35 to col. 8, line 3.	1-3
X	MANNING & NAPIER, DR-LINK: Quick Start Guide, April 1996, Conducting a Search.	1-3

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 SEPTEMBER 1999

Date of mailing of the international search report

08 OCT 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
MICHAEL PEND James R. Matthews
Telephone No. (703) 308-0147

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/17655

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEXIS-NEXIS, Learning Lexis: A Handbook for Modern Legal Research, Nov 1995, page 24.	1-3